Studia Paradyskie · 34/2024
ISSN (print) 0860-8539 · ISSN (online) 2956-4204
DOI: 10.18276/sp.2024.34-18 · s. 301–315

Olgierd Sroczyński

Fundacja Mikołaja Kuzańczyka
olgierd.sroczynski@gmail.com
ORCID: 0000-0003-3437-4097

Zmartwychwstanie przez technikę. Problem tożsamości sztucznej inteligencji w koncepcji transferu umysłu

STRESZCZENIE

Transfer umysłu (ang. *mind uploading*) jest przez autorów kojarzonych z nurtami transhumanistycznymi, wymieniany jako jedna z dróg rozwoju sztucznej inteligencji do poziomu ludzkiego (Artificial General Intelligence). Przekonanie o tym, że tego typu hipotetyczna technologia jest możliwa, bazuje w pierwszej kolejności na redukcjonizmie biologicznym, tj. utożsamieniu umysłu z mózgiem (czy też całym układem nerwowym), po drugie zaś na założeniu obliczalności procesów neurobiologicznych. Spojrzenie na obie przesłanki z punktu widzenia zagadnienia tożsamości osobowej stawia pod znakiem zapytania logiczną możliwość istnienia takiej technologii w sensie, jaki deklarowany jest w transhumanizmie.

SŁOWA KLUCZOWE

transhumanizm, transfer umysłu, dusza, ucieleśnione poznanie, tożsamość

Resurrection through technology. The problem of artificial intelligence identity in the concept of mind uploading

ABSTRACT

Mind uploading is mentioned by authors associated with transhumanist movements as one of the paths to Artificial General Intelligence. The belief that such hypothetical technology is possible to develop, is primarily based on biological reductionism, i.e., equating the mind with the brain (or the entire nervous system), and secondly, on the assumption that neurobiological processes are computable. However, analysing both assumptions from the perspective of personal identity raises doubts about the logical possibility of such technology existing in the way transhumanism proposes.

KEYWORDS

transhumanism, mind uploading, soul, embodied cognition, personal identity

Wprowadzenie

Wraz z bezprecedensowym rozwojem sztucznej inteligencji (SI) w ostatnich latach, który doprowadził do powstania modeli uczenia głębokiego, opartych na technologii transformerów, stare pytania o możliwość stworzenia sztucznej inteligencji na poziomie ludzkim zyskały nową aktualność. Wielkie modele językowe (ang. *Large Language Models*, LLMs), zasilające m.in. roboty konwersacyjne (jak na przykład ChatGPT czy Bard/Gemini), stały się bowiem publiczną sensacją o globalnym zasięgu, ze względu na ich możliwości generowania sensownych odpowiedzi na zadane pytania w niemalże nieograniczonym zakresie tematycznym. Tymczasem zgodnie z funkcjonalistycznymi kryteriami zaproponowanymi przez Alana Turinga fakt, że maszyna licząca udziela odpowiedzi nieodróżnialnych od tych udzielonych przez żywego człowieka, oznacza, że dysponuje ona inteligencją na poziomie równym ludzkiemu[1]. Niezwykła wydajność modeli przetwarzających język naturalny skłania zatem do pojawiających się co jakiś czas twierdzeń, że bariera rozwoju sztucznej inteligencji, to znaczy sztuczna inteligencja ogólna (*Artificial General Intelligence*, AGI), jeżeli jeszcze nie została przekroczona, to zostanie w najbliższym czasie[2].

Zagadnienie maszynowej inteligencji i jej hipotetycznego rozwoju do „poziomu ludzkiego" – przy czym warto zaznaczyć, że definicją takiego poziomu nie dysponujemy – traktowane jest więc jako problem prawie wyłącznie techniczny, oceniany z perspektywy wykazywania pewnych umiejętności, przede wszystkim językowych, na bazie utożsamienia języka z umysłem[3]. Z takiej perspektywy dość łatwo jest przyjmować pewne założenia dotyczące ewentualnego wpływu dalszego rozwoju takiej technologii na społeczeństwo, politykę, gospodarkę czy relacje międzyludzkie. Jednak bardzo często pomija się w tym miejscu – i to programowo – kwestie przynależące do

1 A. Turing, *Computing machinery and intelligence*, „Mind" 59 (1950) 236, s. 433–460.

2 AGI nie ma ustalonej definicji, co też rodzi fundamentalny problem z oceną tego typu prognoz. Najbardziej popularne jest ujęcie funkcjonalistyczne, które określa AGI na podstawie kryteriów dotyczących osiągania przez SI celów w różnych warunkach (np. posługiwanie się wiedzą z różnych dziedzin), co ma odróżniać AGI jako „sztuczną inteligencję silną" (*strong AI*) od „słabej" (*narrow AI*). W taki sposób w kontekście modelu GPT-4 podchodzi do tego zespół Microsoft Research w publikacji mówiącej o przejawach (dosłownie „iskrach") sztucznej ogólnej inteligencji w produkcie firmy OpenAI. Zob. S. Bubeck i in., *Sparks of Artificial General Intelligence: Early experiments with GPT-4*, https://arxiv.org/abs/2303.12712 (30.01.2024). W ujęciu funkcjonalistycznym pomija się kwestię świadomości „silnej" SI, jednak takie wątki w kontekście LLM-ów również się cyklicznie pojawiają. Najgłośniejszy tego typu przypadek jak dotąd stanowi sprawa Blake'a Lemoine'a, inżyniera Google, który ogłosił, że model LaMDA (na którym oparty został później chatbot Bard), zyskał świadomość ludzką. Rewelacje te zostały co prawda od razu zdementowane, niemniej sprawa z tym związana mocno pobudziła publiczną wyobraźnię, co pozwala przewidywać, że również ten temat będzie powracał w debacie na temat AGI. Zob. O. Sroczyński, *Bajki o świadomości robotów*, „Wprost" 28 (2022) 2045, s. 202–219.

3 Por. J. Bremer, M. Flasiński, *The Turing Test, or a misuse of language when ascribing mental qualities to machines*, „Forum Philosophicum" 27 (2022) 1, s. 6–25, https://doi.org/10.35765/forphil.2022.2701.01 (16.03.2024).

„metafizyki umysłu”[4], to znaczy zagadnienie źródeł wolnej woli, świadomości i tożsamości, co stawia pod znakiem zapytania celowość takich rozważań, poza dostarczaniem literackiej rozrywki. W jaki sposób można bowiem sensownie odpowiedzieć na pytanie, co zrobiłaby zaawansowana SI w jakiejkolwiek sytuacji, jeżeli nie jest dla nas jasne chociażby to, czy może ona być źródłem działania w sensie ludzkim. Jeżeli tak, to jakie przesłanki dla jej działania będą istotne? Jeżeli zaś nie, to dlaczego miałaby wyznaczać swoje cele, sprzeczne z celami ludzkimi?

W tym artykule zajmiemy się tymi właśnie zagadnieniami w kontekście jednej z popularniejszych koncepcji drogi do „sztucznej inteligencji na poziomie ludzkim”, to znaczy *mind uploading* (dosłownie „przesyłania” czy „transferu umysłu”, to znaczy zmiany jego postaci na cyfrową). Podstawowe pytanie związane z transferem umysłu dotyczy możliwości odtworzenia konkretnego umysłu ludzkiego na maszynie liczącej – i to umysłu w sensie ścisłym, to jest zdolnego do działania w takim samym stopniu, w jakim zdolny byłby do działania żywy człowiek. Koncepcja cyfryzacji umysłu ludzkiego skupia wymienione problemy jak w soczewce, ponieważ dotyka bezpośrednio zagadnienia ludzkiej tożsamości i świadomości, co pozwala na szersze spojrzenie na „metafizykę umysłu” sztucznej inteligencji, a zatem również na przyszłość jej rozwoju do poziomu ludzkiego.

Modele językowe a cyfrowe kopie umysłu

O ile we wspomnianym wyżej teście Turinga (a także jego późniejszych alternatywach) pytanie o sztuczną inteligencję skupia się na bardzo ogólnej idei wyposażenia komputera w pewien zestaw umiejętności, o tyle wynalazek modeli podstawowych (ang. *foundation models*), przede wszystkim zaś wielkich modeli językowych, przesunął akcenty w tym pytaniu, kierując je na zagadnienie ludzkiej tożsamości. Wielki model językowy (LLM) konstruuje odpowiedzi na podstawie zbioru, na którym został nauczony. Jeżeli więc dysponujemy stosunkowo wąskim zbiorem – na przykład ograniczymy uczenie modelu do zbioru literatury anglojęzycznej z XIX wieku – to odpowiedzi udzielane przez model będą odpowiadały wykrytym w toku uczenia powiązaniom pomiędzy konkretnymi pojęciami, tak samo ograniczona będzie wiedza, jaką dysponuje model. Taki hipotetyczny model zapytany na przykład o mechanikę kwantową lub ogólną teorię względności nie będzie w stanie udzielić sensownej odpowiedzi, ponieważ pojęcia te nie funkcjonują w zbiorze. Będzie natomiast w stanie wygenerować sensowny blok tekstu napisany w stylu Charlesa Dickensa lub Edgara Allana Poego i odpowiedzieć na pytania zgodne ze stanem wiedzy, który z ich dzieł można wnioskować.

4 Zob. N. Bostrom, *Superinteligencja: scenariusze, strategie, zagrożenia*, tłum. D. Konowrocka-Sawa, Warszawa 2016, s. 45.

Modele uniwersalnego zastosowania, takie jak GPT-3.5, GPT-4 czy Gemini, wytrenowane na wielkich zbiorach danych tekstowych, również są w stanie wygenerować teksty stylizowane na konkretnych autorów. Przykładem niech będzie fragment „wiersza o wielkich modelach językowych napisany w stylu Tadeusza Micińskiego", wygenerowany przeze mnie na potrzeby tego artykułu za pomocą modelu GPT-3.5 (ChatGPT):

Wymowność języków, jak pieśń strumieni milczących,
Gdzie dźwięki kropli w płynnym tańcu snują opowieść.
Wielkie modele, jak obłoki, wznoszą się ku gwiazdom,
Gdzie myśli unoszą się jak orły, swobodne i dzikie.

W wielkich modelach, kodów labirynt,
Gdzie myśl się czai, a mowa drzemiąca,
W sieciach neuronów ukryte przesłanie,
Wężowy szept algorytmów nieznanie.

Nie ma znaczenia poetycka wartość wiersza; w tym miejscu należy jednak zwrócić uwagę na odtworzenie bardzo charakterystycznych dla młodopolskiego poety środków poetyckiego wyrazu do opisu zjawiska, które nie mieściło się w ramach pojęciowych epoki Micińskiego. Podobnie zaskakująco trafne wyniki można uzyskać przy generowaniu tekstów stylizowanych na Williama Szekspira, Homera, Johanna W. Goethego czy Williama Blake'a. Mamy więc do czynienia z treścią, która nie jest prostą pochodną istniejących utworów literackich, ale jej adaptacją do opisu zjawisk, które przez oryginalnych autorów nie były znane i niejako spojrzenie na nie z ich perspektywy. Przyglądając się tym i podobnym przykładom, należy zadać sobie pytanie o relację pomiędzy tymi tworami językowymi a przywołanymi autorami. Czy Tadeusz Miciński jest autorem wiersza o wielkich modelach językowych? Jeżeli tak – to w jaki sposób, a jeżeli nie – to kto jest jego autorem: programista, który zaprojektował model, czy osoba wydająca polecenie (*prompt*) modelowi? Pytania te nie są trywialne, budząc przede wszystkim wątpliwości na gruncie prawa autorskiego[5]. Ale odpowiedź na te pytania zależy od kwestii znacznie głębszych: pytania o sens ludzkiej twórczości, o relację pomiędzy językiem a umysłem, wreszcie zaś pytania o istotę ludzkiej indywidualnej tożsamości i jej związek ze świadomością.

5 Pod koniec 2023 r. grupa amerykańskich autorów złożyła pozew przeciwko OpenAI dotyczący naruszenia praw autorskich przez firmę podczas trenowania modelu GPT-3.5; zob. https://www.nytimes.com/2023/09/20/books/authors-openai-lawsuit-chatgpt-copyright.html (30.01.2024). Pytania o to, czy wykorzystanie utworów w procesie uczenia maszynowego jest uprawnione, a także jak traktować produkty modeli na tych utworach trenowanych, są już od kilku lat szeroko dyskutowane i jak dotąd nie został w tej sprawie osiągnięty konsensus. Por. X. Wang, *AI output: A human condition that should not be protected now, or maybe ever*, „Chicago-Kent Journal of Intellectual Property" 20 (2021) 136, https://scholarship.kentlaw.iit.edu/ckjip/vol20/iss1/7 (3.02.2024).

Przykład ten jest użyteczny chociażby w kontekście pojawiających się już od pewnego czasu pomysłów związanych z wykorzystaniem osiągnięć SI do „rozmowy ze zmarłymi bliskimi". Na rynku dostępne są obecnie aplikacje, za pomocą których żywa osoba może nagrać swoje wspomnienia, a następnie – korzystając z modelu językowego generującego wypowiedzi, a także generatora mowy wykorzystującego jej głos – udostępnić je bliskim osobom do wykorzystania po własnej śmierci. Takie technologie wykorzystują obecnie na przykład aplikacje HereAfter AI czy StoryFile[6]. Rozwiązaniom takim jest jeszcze daleko do tego, aby były traktowane przez użytkowników jako „wcielenia" zmarłych osób: ich deklarowanym celem jest m.in. ułatwienie użytkownikom żałoby. „Cyfrowy awatar" stworzony w taki sposób jest po prostu archiwum, co do zasady podobnym do albumu zdjęć, listów czy nagrań wideo pozostawionych przez zmarłego bez udziału SI, jedynie łatwiej obsługiwanym. Rozwiązania te stanowią jednak podstawę do rozważań na temat przyszłości zastosowania tego typu rozwiązań, ponieważ w miarę ulepszania technologii mogą otworzyć możliwość do interpretowania „cyfrowych awatarów" jako rzeczywistych kopii ludzkich umysłów, co z kolei wykracza daleko poza obszar „kontaktu ze zmarłymi". Rodzi to poważne wątpliwości zwłaszcza w kontekście prądu intelektualnego, który zdominował narrację związaną z rozwojem techniki w ogóle, a sztucznej inteligencji w szczególności, tj. transhumanizmu.

Dualizm i wyzwolenie z materii

Pojęcie transhumanizmu obejmuje wiele odmian ideologii, której głównym założeniem jest przekroczenie biologicznych barier życia ludzkiego, co miałoby oznaczać wkroczenie na kolejny etap ewolucji człowieka jako gatunku[7]. Dotyczy to wydłużania życia czy też rozszerzenia percepcji i możliwości ludzkiego mózgu za pomocą środków naukowo-technicznych, w tym implantów, interfejsów łączących mózg z komputerem czy też manipulację kodem genetycznym człowieka. Nie ma tutaj miejsca na to, aby wyczerpująco omówić ideologię transhumanizmu[8], ale należy zwrócić uwagę na wyraźną religijną konotację tej podstawowej deklaracji ideowej, a także na pokrewieństwo, jakie możemy dostrzec pomiędzy transhumanizmem a ideami stricte religijnymi – np. teologią Teilharda de Chardin[9] czy gnostycyzmem[10]. Wskazanie tego pokrewieństwa

6 R. Carballo, *Using A.I. to talk to the dead*, „New York Times" 11.12.2023, https://www.nytimes.com/2023/12/11/technology/ai-chatbots-dead-relatives.html (30.01.2024).

7 J. Huxley, *Transhumanizm*, tłum. M. Soniewicka, „Ethics in Progress" 6 (2015) 1, s. 17–22.

8 Na ten temat zob. M. Garbowski, *Transhumanizm. Geneza – założenia – krytyka*, „Ethos" 28 (2015) 3, s. 23–41.

9 E. Steinhart, *Teilhard de Chardin and transhumanism*, „Journal of Evolution and Technology" 20 (2008) 1, s. 1–22.

10 A. Płoszczyniec, *Transhumanizm jako gnoza. Refleksja analityczno-historyczna*, „Argument" 13 (2023) 1, s. 43–58.

nakierowuje nas na bardzo określoną antropologię, na jakiej transhumanizm się opiera – to zaś ważne jest w kontekście wyżej zadanych pytań.

Gnostycka koncepcja człowieka zakładała, że człowiek – istota boska – jest uwięziony w materialnej powłoce. Świat duchowy i świat materialny były widziane jako całkowicie odmienne i rozłączne, i dualizm ten znajduje odzwierciedlenie w człowieku. Nie była to koncepcja nowa, jako że dualizm – tak antropologiczny, jak i kosmologiczny – znany był już we wcześniejszych religiach (jak irański zoroastryzm), znalazł także swoje miejsce w filozofii starożytnej, w tym w najbardziej wpływowej – platońskiej. Co jednak odróżniało gnostycyzm od wcześniejszych wersji dualizmu, to jego radykalny antykosmicyzm i antymaterializm[11]. Odrzucenie ciała prowadziło do bardzo określonego stylu życia sekt gnostyckich, jak bogomiłowie czy katarzy, z wysuwającymi się na pierwszy plan skrajnym ascetyzmem i powstrzymaniem się od prokreacji jako wyrazem niezgody na uczestnictwie w grzesznym dziele Demiurga-Szatana. Celem konsekwentnego gnostycyzmu byłoby więc wymarcie ludzkości[12].

Jednak trzymając się gnostyckiego dualizmu, można położyć nacisk nie na odrzucenie świata, ale na boskość człowieka uwięzionego w materii. Takie właśnie są korzenie „gnozy nowożytnej", jak charakteryzował ją Eric Voegelin. Przedstawiciele tej gnozy nie odwołują się do Walentyna czy Marcjona, nie wierzą w diabelskie pochodzenie materii, w grzech i w zbawienie człowieka po śmierci. Podstawą ich myślenia jest ubóstwienie nauk przyrodniczych: nauk, które ogołociły człowieka i świat z tajemnicy, ale – jako twór ludzkiego umysłu – dają nadzieję na przekształcenie materialnego świata w taki sposób, aby wyzwolić ludzkość z ograniczeń biologicznych[13]. Elementy tego stanowiska znajdujemy między innymi w utopizmie XVI i XVII wieku, w oświeceniowym i marksistowskim materializmie, w mitologii narodowego socjalizmu, wreszcie we współczesnym transhumanizmie. We wszystkich wymienionych ideologiach odnaleźć możemy dążenie do „immanentyzacji eschatonu", czyli realizacji celów ostatecznych człowieka w ziemskiej rzeczywistości za pomocą doczesnych środków. Wszystko, co wykracza poza ramy „naukowego", czyli materialistycznego opisu świata, jest w nich odrzucone[14].

---

11 K. Rudolph, *Gnoza*, tłum. G. Sowiński, Kraków 2011, s. 67; H. Jonas, *Religia gnozy*, tłum. M. Klimowicz, Kraków 1994, s. 266–270.

12 Ch. Godin, *Koniec ludzkości*, tłum. Z. Pająk, Kraków 2004, s. 123–124.

13 „Wielu ludzi obecnie podziela gnostycki ogląd rzeczy, nie zdając sobie z tego sprawy. Wierząc, że człowiek może być w pełni wyjaśniony za pomocą naukowego materializmu, odrzucają ideę wolnej woli. Jednocześnie jednak nie są w stanie porzucić nadziei na bycie panami własnego przeznaczenia. Muszą więc wierzyć, że nauka w jakiś sposób umożliwi ludzkiemu umysłowi ucieczkę z ograniczeń, jakie natura nań nakłada" – J. Gray, *The soul of the marionette. A short inquiry into human freedom*, New York 2016.

14 „Nowożytność to gnoza radykalnie zsekularyzowana, gdzie mistyczne i ezoteryczne nauki zastępuje nauka, kult człowieka i jego możliwości. Jest to więc totalne załamanie hierarchii dokonane przez immanentystów. W zsekularyzowanej gnozie człowiek sam zaczyna uważać się za Boga, stąd też pojęcie »egofanii«, który poprzez rozum, naukę i technikę kreuje lepszy świat" – A. Wielomski, *Historia filozofii w ujęciu Erica Voegelina*, „Kwartalnik Historii Nauki i Techniki" 44 (1999) 1, s. 52–53.

Wszelako dążenie do przekroczenia granic biologii, które jest kluczowe dla transhumanizmu, zawiera w sobie gnostycką pogardę dla ciała i stanowi punkt wyjścia dla paradoksalnego dualizmu. Odrzucając duchowość, transhumaniści uznają umysł za emergentną własność materialnego ciała ludzkiego. Zakłada się jednak, że wyczerpujący jego opis – jak na razie hipotetyczny – powstały na przykład w wyniku skanowania mózgu, może pozwolić na odtworzenie struktury połączeń neuronalnych w programie komputerowym. Pozwoliłoby to więc na przeniesienie ludzkiej tożsamości do postaci cyfrowej[15]. Umysł ludzki zatem stałby się wówczas niematerialny i właściwie niezależny od nośnika, na którym w danym momencie się znajduje: jeżeli bowiem byłby w stanie zachować tożsamość po przeniesieniu z ciała na konkretny komputer, to zachowałby ją również skopiowany na inne komputery, a także przy nabyciu nowych zasobów pamięci i mocy obliczeniowej. „Kolejny etap ewolucji człowieka", czyli transfer umysłu, oznaczałby więc według transhumanistycznej ideologii bytowanie *de facto* duchowe, w którym ograniczenia biologiczne przestałyby mieć znaczenie[16].

Umysł bez ciała

Nie można wykluczyć obecnie, zwłaszcza jeżeli zważyć na całkiem niedawne postępy w tym zakresie[17], że wykorzystanie skanowania mózgu pozwoli na stworzenie bardziej przekonujących cyfrowych kopii ludzkiego umysłu, niż pozwalają na to obecnie technologie obróbki języka naturalnego w uczeniu głębokim. Jeżeli skupić się na przywołanym na początku zagadnieniu wykorzystania takich kopii do „kontaktów ze zmarłymi", to konsekwencją takiego przełomu byłaby dalsza eksploatacja pogrążonych w żałobie osób. To zjawisko pozostaje jednak w obszarze interakcji człowieka z programem komputerowym, nie ma więc większego znaczenia, czym *naprawdę* program jest – zgodnie z funkcjonalistycznymi kryteriami wystarczy przekonanie jego żywego użytkownika, że program mówiący głosem bliskiego zmarłego jest naprawdę jakąś jego wersją. To, co nas interesuje w tym miejscu, to na ile hipotetyczna kopia cyfrowa umysłu ludzkiego ma szansę rzeczywiście *być* umysłem ludzkim. Trzymając się przykładu przywołanego na początku: na ile model stworzony na bazie skanowania mózgu

15 S. Cave, *AI: Artificial Immortality and narratives of mind uploading*, [w:] *AI Narratives: A history of imaginative thinking about intelligent machines*, red. S. Cave, K. Dihal, S. Dillon, Oxford 2020, s. 309–332.

16 R. Kurzweil, *Singularity is near*, New York 2005, s. 38.

17 Chodzi tutaj przede wszystkim o badania prowadzone przez firmę Neuralink, które w marcu 2024 r., zaledwie kilka miesięcy po rozpoczęciu testów na ludziach, zaowocowały wykorzystaniem implantu mózgowego u sparaliżowanego w wyniku wypadku pacjenta do czynności takich, jak gra w szachy czy pisanie. Zob. https://www.ndtv.com/world-news/man-with-elon-musks-neuralink-brain-chip-posts-tweet-on-x-just-by-thinking-5300011 (24.03.2024).

Tadeusza Micińskiego (o ile byłoby to możliwe) byłby „bardziej" Tadeuszem Micińskim, niż model językowy wytrenowany na pismach młodopolskiego poety?

Założenie o realnej możliwości *mind uploading*, jakie stanowi podstawę transhumanistycznej wiary w przyszłość gatunku ludzkiego, związane jest bezpośrednio z komputacjonizmem, nurtem filozofii umysłu, według którego czynności ludzkiego umysłu można wyczerpująco opisać za pomocą operacji obliczeniowych[18]. Komputacjonizm garściami czerpie z Kartezjańskiego dualizmu, skupiając się na języku, jako wyrazie ludzkiego myślenia, oraz na działaniu mózgu, jako maszyny do jego przetwarzania. Jeżeli zredukujemy myślenie ludzkie do przetwarzania symboli według określonych reguł, to konieczny staje się wniosek o tym, że odtworzenie ludzkiego mózgu na maszynie liczącej jest co do zasady możliwe[19]. Pojawia się jednak fundamentalny problem.

Przyjmijmy hipotetycznie, że posiadamy szczegółową mapę mózgu osoby *A*, zawierającą nie tylko wszystkie należące do niej połączenia neuronalne, ale również wykształcone przez nią reguły przetwarzania informacji. Wiemy zatem, jakie neurony odpalane są w reakcji na konkretny bodziec i przy przetwarzaniu jakiej informacji. Posługując się tym opisem, tworzymy model mózgu *A'* i uruchamiamy go na komputerze o dużej mocy obliczeniowej. Test Turinga powinien – według konsekwentnego komputacjonizmu – ujawnić taki sam *output* w odpowiedziach *A* i *A'*, jeśli bylibyśmy w stanie wyizolować *A* ze wszystkich bodźców, których nie doświadcza *A'*, lub też zasymulować w *A'* wszystkie bodźce, którym poddany jest *A*.

Gdyby nawet tak złożony model obliczeniowy był możliwy, w takim eksperymencie od razu napotykamy trudność związaną z warunkiem odizolowania *A*, czyli żywego człowieka, od bodźców, których nie doświadcza jego model. Podobnie niemożliwe jest – przy dysponowaniu samym skanem mózgu – zasymulowanie w *A'* wszystkich informacji, jakie odbiera mózg *A*. Ciało *A* przez cały czas bowiem odbiera bodźce ze środowiska, ale również z wnętrza własnego organizmu, w którym nieprzerwanie toczą się procesy biologiczne[20]. Zatem już na wstępie rozważań nie mówimy o symulacji mózgu, ale o symulacji całego ciała – w każdym innym wypadku *A'* nie byłby pełnym i wyczerpującym modelem *A*.

Nie jest to mocny argument przeciwko komputacyjnej teorii umysłu[21], a zwłaszcza przeciw nadbudowanej na niej transhumanistycznej fantazji: bazując na Kartezjańskim dualizmie moglibyśmy po prostu zadeklarować, że w takim modelowaniu mamy

18 W przypadku uczenia głębokiego i modelowania umysłu ludzkiego za pomocą sieci neuronowych mówimy o koneksjonizmie. Koneksjonizm nie jest jednak odrębnym od komputacjonizmu nurtem kognitywistyki, ale jego szczególną odmianą, zob. S. Pinker, *Jak działa umysł*, tłum. M. Koraszewska, Poznań 2022.

19 Por. m.in.: N. Bostrom, *The future of human evolution*, [w:] *Death and anti-death: Two hundred years after Kant, fifty years after Turing*, red. C. Tandy, Ann Arbor 2004.

20 Por. N. Herzfeld, *The artifice of intelligence: Divine and human relationships in a robotic age*, Minneapolis 2023, s. 35.

21 Dokonuję tutaj pewnego uproszczenia idei stojącej za komputacjonizmem/koneksjonizmem. Szersze omówienie koneksjonizmu można znaleźć m.in. w: I.S.N. Berkeley, *The curious case of connectionism*,

do czynienia z „czystą" myślą, która nie jest zaburzona przez „zwierzęce instynkty"[22]. Umysł bowiem jest emergentny wobec ciała, ale nie jest sumą składową wszystkich jego elementów[23]. W tej perspektywie *A'* byłby nie tyle kopią, co ulepszoną wersją *A*, ponieważ pozbawioną „szumu informacyjnego" produkowanego przez ludzkie ciało i środowisko. Taki model mógłby więc zostać uznany za „duszę" *A*, boski pierwiastek uwolniony z ograniczeń materialnego ciała, co pozostaje w zgodzie z quasi-religijną deklaracją ideową transhumanizmu.

Jednak pytanie, jakie nieuchronnie w tym miejscu się pojawia, dotyczy wpływu takiego „uwolnienia" na działanie umysłu/duszy, a zatem również na użyteczność takiego modelu, skoro tylko wyjdziemy poza możliwe użycie go jako „cyfrowego awataru/magazynu informacji" dostępnego po śmierci. Taki przypadek użycia, chociaż bez wątpienia budzący kontrowersje moralne, jest stosunkowo łatwy do konceptualizacji. Inaczej ma się sprawa z traktowaniem „cyfrowego umysłu" jako osoby ludzkiej, co jest celem właściwym *mind uploading*. Nawet jeżeli dodamy do naszej definicji wspomniane już transhumanistyczne zastrzeżenie, że rozmawiamy o przyszłym kształcie ludzkości, „dalszym etapie ewolucji" – a więc nie możemy tego oceniać w kategoriach ludzkich – to problematyczny z punktu widzenia logiki pozostaje sens takiego działania.

Rola ciała w perspektywie antropologii chrześcijańskiej

Analizując to zagadnienie, wypada w pierwszej kolejności odwołać się do spojrzenia teologicznego, jako że eschatologia transhumanistyczna wyraźnie wkracza na domenę teologii – nie tylko w pojawiających się stosunkowo często próbach łączenia jej z chrześcijańskimi koncepcjami duszy i zmartwychwstania[24], ale również funkcjonalnie, jako substytut wiary[25].

W antropologii biblijnej podział na duszę i ciało istnieje, ale nie w sensie, jaki obecny jest w późniejszych koncepcjach dualistycznych. Człowiek pojmowany jest jako nierozdzielna całość, istota materialna napełniona tchnieniem Bożym[26]. Według E. Voegelina

„Open Philosophy" 2 (2019), s. 190–205. Por. także: J. Bremer, *Wprowadzenie do filozofii umysłu*, Kraków 2010, s. 136–148.

22 D.F. Noble, *Religia techniki*, tłum. K. Kornas, Kraków 2017, s. 212–213.

23 M. Hauskeller, *My brain, my mind, and I. Some philosophical assumptions of mind- uploading*, „International Journal of Machine Consciousness" 4 (2012) 1, s. 187–200.

24 D.F. Noble, *Religia techniki*, s. 234–253.

25 Zob. M. Laakasuo i in., *Would you exchange your soul for immortality? – existential meaning and afterlife beliefs predict mind upload approval*, „Frontiers in Psychology" 14 (2023), https://doi.org/10.3389/fpsyg.2023.1254846 (2.03.2024). Por. O. Sroczyński, *Teologiczny sens mitów o sztucznej inteligencji*, „Kieleckie Studia Teologiczne" 22 (2023), s. 113–127.

26 A. Świderkówna, *Zmartwychwstanie – jak zrozumieć tajemnicę?*, [w:] tejże, *Nie tylko o Biblii*, Warszawa 2005, s. 54.

takie pojmowanie człowieka było udziałem całego świata starożytnego, aż do momentu, w którym filozofia grecka odkryła człowieka na nowo i doprowadziła do jego podziału na ciało i duszę[27]. Z kolei teologia chrześcijańska, czerpiąc z obu tradycji, godziła te stanowiska, przyjmując Arystotelesowski hylemorfizm, z podziałem na formę i materię, akt i możność. Osoba ludzka to *synolon*, nierozdzielne złożenie duszy jako formy materii, ponieważ dusza ludzka – jako najniższa w hierarchii bytów duchowych – potrzebuje ciała, aby móc poznawać:

> A doskonałość wszechświata wymagała tego, żeby rzeczy były ułożone według różnych stopni. Gdyby zatem Bóg tak urządził ludzkie dusze, żeby poznawały w sposób przysługującym jestestwom istniejącym bez materii, nie miałyby poznania doskonałego, a tylko niewyraziste, mgliste i ogólnikowe. Na to więc, żeby mogły mieć doskonałe i właściwe poznanie rzeczy zostały z natury tak urządzone, żeby łączyły się z ciałami i żeby przez to z samych rzeczy postrzegalnych czerpały właściwe o nich poznanie; podobnie jak z ludźmi prostymi: nie można ich doprowadzić do wiedzy inaczej, jak tylko za pomocą postrzegalnych przykładów uzmysławiających rzeczy[28].

Święty Tomasz nie neguje możliwości poznania pozazmysłowego, jednak poznanie to nie dotyczy już działania w świecie materialnym. *Anima separata* – dusza oddzielona od ciała – nie poznaje już świata materialnego i nie nabywa o nim żadnych nowych informacji. Posiada za to pogłębiony wgląd w siebie, w informacje już nabyte, bez odwoływania się do wyobrażeń – co w sytuacji połączenia z ciałem nie było możliwe[29]. Chociaż dusza ludzka, jako nieśmiertelna forma i akt materialnego ciała, jest w antropologii św. Tomasza źródłem ludzkiej godności, to wciąż samodzielnie nie posiada statusu ontologicznego osoby ludzkiej.

Wspomniane przez św. Tomasza odwoływanie się do wyobrażeń ma szczególne znaczenie w kontekście ludzkiej tożsamości. Chociaż bowiem językowo odwołujemy się do pojęć ogólnych – i bez takiej obiektywizacji nie miałaby racji bytu żadna komunikacja – to jednak proces *rozumienia* poszczególnych pojęć jest ściśle zindywidualizowany. Każdy kompetentny użytkownik języka będzie w stanie zrozumieć pojęcia takie, jak „ojciec", „pies", „miasto" użyte w dowolnym sensownym zdaniu, ale na hasło „wyobraź sobie ojca/psa/miasto" każdy będzie odwoływał się do zupełnie innych elementów doświadczenia zmysłowego, które z kolei będą miały wpływ na ich rozumienie, a w efekcie na cały następujący po takim ćwiczeniu proces mentalny. Wszelkie doświadczenia zmysłowe

27 Por. A. Wielomski, *Historia filozofii*, s. 43–48.

28 Tomasz z Akwinu, *Suma teologiczna*, t. 7, cz. II, q. 89, tłum. P. Bełch OP, https://www.katedra.uksw.edu.pl/suma/suma7.htm (28.02.2024).

29 P. Mrzygłód, *Filozoficzne rozumienie śmierci oraz jej interpretacja w świetle antropologii św. Tomasza z Akwinu*, „Perspectiva. Legnickie Studia Teologiczno-Historyczne" 1 (2011) 18, s. 153–157.

mają bowiem bardzo określony kontekst, nie tylko środowiskowy, ale również biologiczny, w tym emocjonalny[30].

To „ucieleśnione poznanie" stanowi podstawę dla koncepcji indywidualnej tożsamości i niepowtarzalności ludzkich osób. Rozważany przez nas hipotetyczny model *A'* nie musi być traktowany jako zamknięty zbiór wiedzy i przekonań *A* i może podlegać dalszemu uczeniu. Ale uczenie takie będzie się odbywało już poza kontekstem ciała *A* – zatem w takim scenariuszu będziemy mieli do czynienia z wyjściem poza tożsamość pierwowzoru modelu. Innymi słowy wszystko to, co będzie reprezentował model *A'*, mogłoby być reprezentowane przez całkowicie inny model, oparty na neuronalnej strukturze zupełnie innej osoby, ponieważ z konieczności proces uczenia musiałby być zobiektywizowany i odcieleśniony. Model *A'* nie zdobywałby już wiedzy jako kopia osoby *A*, ale jako „jakiś" – pozbawiony tożsamości – model.

Wolna wola

W świetle tych przesłanek wszystkie wypowiedzi i ewentualne działania modelu *A'* byłyby zatem pochodną procesów mentalnych *A*, ale nie mogłyby być traktowane jako substytut wypowiedzi i działań *A*. Obserwacja ta ma fundamentalne znaczenie nie tylko dla rozważanej przez nas kwestii możliwości swoistego „zmartwychwstania" dzięki sztucznej inteligencji, do czego sprowadza się koncepcja *mind uploading*. Człowiek, którego umysł zostałby przeniesiony w środowisko cyfrowe, nie zyskałby przez to „drugiego życia", ponieważ pozbawiony zostałby wszystkich narzędzi umożliwiających mu aktywne rozumienie tego, co wokół niego się dzieje. Jego model mógłby na przykład udzielać sensownych wypowiedzi na temat zdarzeń, które miały miejsce już po jego śmierci, ale nie mamy podstaw, aby sądzić, że byłyby to wypowiedzi, których udzieliłby on sam. Byłby on ich autorem w takim samym stopniu, w jakim Tadeusz Miciński jest autorem cytowanego wiersza o wielkich modelach językowych, albo Claude Cahun jest autorem obrazów wygenerowanych przez popularny model Midjourney przy użyciu polecenia zawierającego frazę „by Claude Cahun"[31].

30 Odkrycie wpływu czynników biologicznych, w tym właśnie emocji, na funkcjonowanie ludzkiego umysłu, rozwijane obecnie w nurcie kognitywistyki nazywanym „ucieleśnionym poznaniem" (*embodied cognition*), zawdzięczamy przede wszystkim pracom neurobiologa A. Damasio oraz językoznawców G. Lakoffa i M. Johnsona. Zob. A. Damasio, *Descartes' error: emotion, reason, and the human brain*, New York 1994; G. Lakoff, M. Johnson, *Metaphors we live by*, Chicago 1980; tychże, *Philosophy in the flesh: The embodied mind and its challenge to western thought*, New York 1999. Przystępne wprowadzenie do koncepcji ucieleśnionego poznania można znaleźć również w książce B.K. Bergena, *Latające świnie. Jak umysł tworzy znaczenie*, tłum. Z. Lamża, Kraków 2018.

31 Przykład zaczerpnięty z rankingu najpopularniejszych stylów używanych przy generowaniu obrazów przez Midjourney przy użyciu biblioteki Midlibrary: https://midlibrary.io/tops/most-collected-styles (15.04.2024).

W ten sposób dotykamy zagadnienia głębszego, związanego z działaniami, jakie podejmowałaby hipotetyczna sztuczna inteligencja na poziomie ludzkim. W scenariuszach narodzin „superinteligencji", to znaczy sztucznej ogólnej inteligencji, która dzięki nieograniczonym możliwościom uczenia staje się znacznie bardziej inteligentna od przeciętnego człowieka, i wskutek tego zaczyna zagrażać ludziom, SI ma możliwość podejmowania własnych decyzji co do celów, a już na pewno środków do ich realizacji. Obrazowy jest przykład „paperclip maximizer", używany przez filozofa-transhumanistę Nicka Bostroma: nawet jeżeli zdefiniujemy superinteligencji bardzo wąski cel, na przykład produkcję spinaczy do papieru, może ona w toku jego realizacji zacząć pozyskiwać środki do produkcji w sposób, który będzie stał w sprzeczności z egzystencjalnym bezpieczeństwem ludzkości[32]. Przykład ten jest wewnętrznie sprzeczny, ponieważ z jednej strony jesteśmy w stanie zdefiniować superinteligencji cel – nie posiada ona zatem możliwości samodzielnego decydowania o tym, co będzie robiła – z drugiej strony zaś dopuszczamy możliwość, że może ona nie być zdeterminowana w jego realizacji.

Przykładu Bostroma można bronić jako użytej w celach retorycznych hiperboli, która ma za zadanie zwrócić uwagę na nieprzewidywalne skutki działania zaawansowanych technologii. To, że dana maszyna działa deterministycznie, nie oznacza, że jesteśmy w stanie w dowolnym momencie przewidzieć jej przyszłe położenie. Efekt działania zaawansowanego modelu algorytmicznego może być więc „zaskakujący" dla ludzkiego obserwatora; kwestia ta została zresztą omówiona już przez Alana Turinga[33]. Nie ma sprzeczności między determinizmem a nieprzewidywalnością, ponieważ oba elementy mieszczą się w ramach matematycznego opisu. Jednak ta kwestia dotyczy dokładnie każdej maszyny, włączając w to najbardziej proste. Fakt, że szlifierka kątowa jest w stanie odciąć mi palec, odbijając się w sposób nieprzewidywalny od powierzchni, na której pracuję, nie oznacza wszakże, że „zrealizowała cel", który był sprzeczny z moim. Skoro nie przypisujemy takim maszynom intencji, to czemu inaczej mielibyśmy traktować zaawansowane modele algorytmiczne? Wymagałoby udowodnienia, że istnieje fundamentalna różnica pomiędzy maszynami liczącymi a innymi maszynami, umożliwiająca traktowanie tych pierwszych jako źródła działania w sensie ludzkim, albo też przyjęcia radykalnego stanowiska deterministycznego, odrzucającego jakiekolwiek rozróżnienia w tym zakresie.

Nawet jeżeli chcielibyśmy zanegować istnienie wolnej woli u człowieka, opierając się na materialistycznych wersjach inkompatybilizmu i traktując poczucie wolności decyzji jako generowaną przez mózg iluzję[34], nie uzasadnia to w żaden sposób możliwości

32 N. Bostrom, *Superinteligencja*, s. 162.

33 S. Delacroix, *Computing machinery, surprise and originality*, „Philosophy & Technology" 34 (2021), s. 1195–1211, https://doi.org/10.1007/s13347-021-00453-8 (28.02.2024).

34 Por. J. Bremer, *Czy wolna wola jest wolna? Kompatybilizm na tle badań interdyscyplinarnych*, Kraków 2010, s. 37–40.

zaistnienia takiego – lub choćby zbliżonego – fenomenu w „umyśle" sztucznej inteligencji (ani w żadnym innym od ludzkiego). Patrząc na człowieka wyłącznie z perspektywy materializmu, zmuszeni jesteśmy powiązać istnienie fenomenu wolnej woli, czyli przede wszystkim poczucia sprawstwa, z uświadomionym pojęciem indywidualnej tożsamości – a to z kolei prowadzi nas nieuchronnie do kondycji człowieka jako istoty ucieleśnionej i biologicznej, ograniczonej czasowo i przestrzennie. Nie ma znaczenia, czy tożsamość i wolna wola są iluzjami, czy nie. Faktem jest natomiast, że wpływ na ludzkie postrzeganie świata, łączenie obserwowanych zdarzeń w łańcuchy przyczynowo-skutkowe, a więc nadawanie sensu rzeczywistości i własnym działaniom, ma bycie *tym konkretnym* organizmem, ograniczonym przestrzennie i czasowo, śmiertelnym i motywowanym biologicznymi popędami.

Ujmując wolną wolę wyłącznie fenomenalistycznie, możemy więc dostrzec, że jest ona – paradoksalnie – ściśle zależna od ludzkiej biologii. Jednak wychodząc poza fenomen, dostrzegamy wciąż tę samą zależność: chociaż tworzenie sensu odbywa się na podstawie danych zmysłowych, to ludzki umysł z łatwością poza nie wychodzi, organizując rzeczywistość wedle sobie tylko właściwych i niewystępujących w przyrodzie obrazów, w tym obrazu samego siebie. Tę zdolność możemy uznać za wyróżnik człowieczeństwa[35]. Dzięki tej zdolności ludzki umysł nie jest biernym rejestratorem zdarzeń, którym następnie przypisuje sens, ale aktywnym organizatorem poznania: to, na czym skupiamy uwagę w danym momencie, które fakty rejestrujemy i w jaki sposób je następnie interpretujemy, jest w dużej mierze kwestią wolnej decyzji poznawczej, opartej właśnie na zdolności tworzenia obrazów. Same zdolności obliczeniowe nie wyjaśniają potrzeby tworzenia obrazów i narracji sensotwórczych, niezależnie od tego, czy chodzi o sztukę naśladowczą, fikcję czy naukę. Tego typu aktywność ludzkiego umysłu jest fundamentalnie nieredukowalna do obliczeń[36].

Skoro sztuczna inteligencja – niezależnie od tego, czy będzie to skonstruowany model wytrenowany na określonym zbiorze danych, czy też skan mózgu konkretnej osoby – ma być pozbawiona biologicznych „ograniczeń", jakie właściwe są gatunkowi ludzkiemu, to nie ma podstaw do przypuszczenia, że będzie również dysponowała samoświadomością, a więc i poczuciem sprawstwa. Mogłaby więc wykonać każde dowolne zadanie – ale nie mogłaby *chcieć* go wykonać. Nie znosi to niebezpieczeństwa związanego z nieprzewidywalnością działania maszyn liczących. To, że nie przypiszemy im moralnej odpowiedzialności za działania, nie wyklucza tego, że mogą faktycznie dokonać szkód, które nie będą intencją ich projektantów. Ale brak wolnej woli nie jest tutaj wyłącznie problemem etycznym, ponieważ zasadniczo wpływa na epistemologię: bez tej właściwości sztuczna inteligencja nie będzie w stanie wejść na poziom rzeczywiście ludzki.

35 H. Jonas, *The phenomenon of life. Toward a philosophical biology*, New York 1966, s. 185–186; tenże, *Homo pictor and the differentia of man*, „Social Research" 29 (1962) 2, s. 201–220.

36 Por. R. Penrose, *Emperor's new mind: Concerning computers, minds, and the laws of physics*, Oxford 1989.

Podsumowanie

Rozważania dotyczące tożsamości i wolnej woli prowadzą do wniosku, że idea *mind uploading* nie ma szansy na realizację, przynajmniej w obecnym paradygmacie i przy obecnym stanie wiedzy. Dotyczy to jednak nie tylko samego kopiowania ludzkiego umysłu do postaci cyfrowej, ale rzutuje na całość założeń dalszego rozwoju SI i ewentualnych zagrożeń z tym związanych. Nie można wykluczyć, że powstanie „sztuczna inteligencja na poziomie ludzkim", jeżeli będziemy ją definiować na podstawie kryteriów określonych w popularnych testach funkcjonalistycznych. Nie ma jednak żadnych podstaw do przypuszczeń, że kiedykolwiek – niezależnie od skomplikowania modeli – będzie ona zdolna do rzeczywiście ludzkiego działania.

Bibliografia

Bergen B.K., *Latające świnie. Jak umysł tworzy znaczenie*, tłum. Z. Lamża, Kraków 2018.

Berkeley I.S.N., *The curious case of connectionism*, „Open Philosophy" 2 (2019), s. 190–205.

Bostrom N., *The future of human evolution*, [w:] *Death and anti-death: Two hundred years after Kant, fifty years after Turing*, red. C. Tandy, Ann Arbor 2004.

Bostrom N., *Superinteligencja: scenariusze, strategie, zagrożenia*, tłum. D. Konowrocka-Sawa, Warszawa 2016.

Bremer J., *Wprowadzenie do filozofii umysłu*, Kraków 2010.

Bremer J., *Czy wolna wola jest wolna? Kompatybilizm na tle badań interdyscyplinarnych*, Kraków 2010.

Bremer J., Flasiński M., *The Turing Test, or a misuse of language when ascribing mental qualities to machines*, „Forum Philosophicum" 27 (2022) 1, s. 6–25, https://doi.org/10.35765/forphil.2022.2701.01.

Bubeck S. i in., *Sparks of Artificial General Intelligence: Early experiments with GPT-4*, Microsoft Research, https://arxiv.org/abs/2303.12712.

Carballo R., *Using A.I. to talk to the dead*, „New York Times" 11.12.2023, https://www.nytimes.com/2023/12/11/technology/ai-chatbots-dead-relatives.html.

Cave S., *AI: Artificial Immortality and narratives of mind uploading*, [w:] *AI Narratives: A history of imaginative thinking about intelligent machines*, red. S. Cave, K. Dihal, S. Dillon, Oxford 2020, s. 309–332.

Damasio A., *Descartes' error: emotion, reason, and the human brain*, New York 1994.

Delacroix S.. *Computing machinery, surprise and originality*, "Philosophy & Technology" 34 (2021), s. 1195–1211, https://doi.org/10.1007/s13347-021-00453-8.

Garbowski M., *Transhumanizm. Geneza – założenia – krytyka*, „Ethos" 28 (2015) 3, s. 23–41.

Godin Ch., *Koniec ludzkości*, tłum. Z. Pająk, Kraków 2004.

Gray J., *The soul of the marionette: A short inquiry into human freedom*, Farrar, Straus and Giroux, New York 2016.

Hauskeller M., *My brain, my mind, and I. Some philosophical assumptions of mind-uploading*, „International Journal of Machine Consciousness" 4 (2012) 1, s. 187–200.

Herzfeld N., *The artifice of intelligence: Divine and human relationships in a robotic age*, Minneapolis 2023.

Huxley J., *Transhumanizm*, tłum. M. Soniewicka, „Ethics in Progress" 6 (2015) 1, s. 17–22.

Jonas H., *Homo pictor and the differentia of man*, „Social Research" 29 (1962) 2, s. 201–220.

Jonas H., *The phenomenon of life. Toward a philosophical biology*, New York 1966.

Jonas H., *Religia gnozy*, tłum. M. Klimowicz, Kraków 1994.

Kurzweil R., *Singularity is near*, New York 2005.

Laakasuo M. i in., *Would you exchange your soul for immortality?– existential meaning and afterlife beliefs predict mind upload approval*, „Frontiers in Psychology" 14 (2023), https://doi.org/10.3389/fpsyg.2023.1254846 (2.03.2024).

Lakoff G., Johnson M., *Metaphors we live by*, Chicago 1980.

Lakoff G., Johnson M, *Philosophy in the flesh: The embodied mind and its challenge to western thought*, New York 1999.

Mrzygłód P., *Filozoficzne rozumienie śmierci oraz jej interpretacja w świetle antropologii św. Tomasza z Akwinu*, „Perspectiva. Legnickie Studia Teologiczno-Historyczne" 1 (2011) 18, s. 141–164.

McCarthy J., *Ascribing mental qualities to machines*, [w:] *Philosophical perspectives in artificial intelligence*, red. M. Ringle, London 1979.

Noble D., *Religia techniki. Boskość człowieka i duch wynalazczości*, tłum. K. Kornas, Kraków 2017.

Pinker S., *Jak działa umysł*, tłum. M. Koraszewska, Poznań 2022.

Płoszczyniec A., *Transhumanizm jako gnoza. Refleksja analityczno-historyczna*, „Argument" 13 (2023) 1, s. 43–58.

Rudolph K., *Gnoza*, tłum. G. Sowiński, Kraków 2011.

Sroczyński O., *Bajki o świadomości robotów*, „Wprost" 28 (2022), s. 202–219.

Sroczyński O., *Teologiczny sens mitów o sztucznej inteligencji*, „Kieleckie Studia Teologiczne" 22 (2023), s. 113–127.

Świderkówna A., *Nie tylko o Biblii*, Warszawa 2005.

Wang X., *AI output: A human condition that should not be protected now, or maybe ever*, „Chicago-Kent Journal of Intellectual Property" 20 (2021) 136, https://scholarship.kentlaw.iit.edu/ckjip/vol20/iss1/7 (3.02.2024).

Wielomski A., *Historia filozofii w ujęciu Erica Voegelina*, „Kwartalnik Historii Nauki i Techniki" 44 (1999) 1, s. 33–63.

CYTOWANIE

O. Sroczyński, *Zmartwychwstanie przez technikę. Problem tożsamości sztucznej inteligencji w koncepcji transferu umysłu*, Studia Paradyskie 34 (2024), s. 301–315, DOI: 10.18276/sp.2024.34-18.